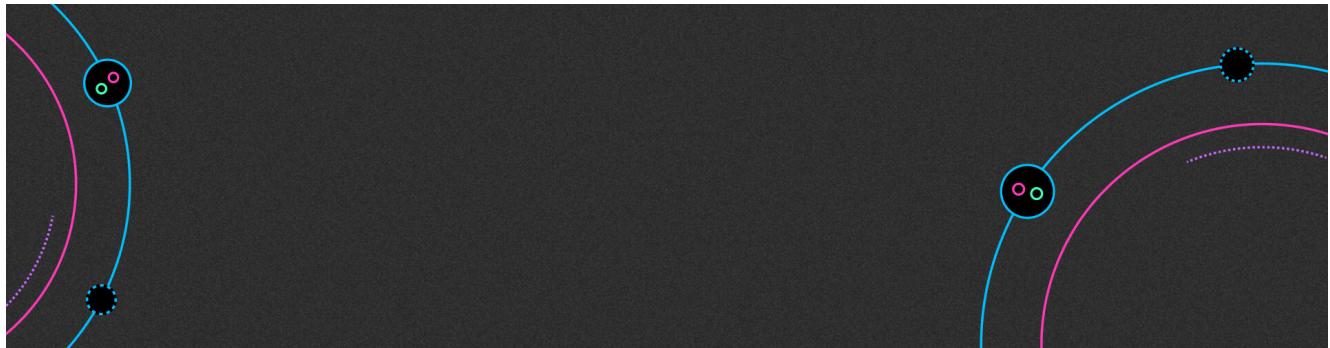


What Builders Talk About When They Talk About AI



The foundation models we use today aren't just beating humans at Jeopardy, or [playing Go](#) at a supernatural level, or powering better recommendation engines—they're changing and creating new user behavior across art, design, games, medicine, coding, and even human relationships. And the broad impact of this wave of AI is already creating the sort of massive value we've seen with other platform shifts, like the internet, mobile, cloud.

Given we're still at the beginning of this paradigm shift, however, it can be tough to envision where these initial breakthroughs will lead and how they'll transform our world. We brought together some of the best AI founders and leaders—from those scaling the underlying foundation models to those building products that could transform entire industries—for a series of conversations on where we are, where we're going, and the big open questions in the field.

Here are the top 16 things we learned.

Mira Murati OpenAI

• **David Baszucki** Roblox

Dylan Field Figma

• **Ali Ghodsi** Databricks

Kevin Scott Microsoft

• **Daphne Koller** insitro

Dario Amodei Anthropic

• **Noam Shazeer** Character.AI



 AI ÷ a16z Builders

Full conversations on a16z.com/AIRevolution and [YouTube](#).

AI today

1. We're at the beginning of the third epoch of compute

The most common refrain across those building AI is that AI is emerging as the next computing platform, and we are still very, very early in that process. Previous platform shifts show it can take decades of compounding innovations for a technological breakthrough to turn into a society-wide shift. The light bulb was invented in 1879, but it wasn't until 1930 that 70% of US houses had electricity. The microchip was invented in 1958, but it took

almost 50 years for us to get the iPhone.

While a lot of the focus today is on the development of foundational large language models (LLMs), the transformer architecture was invented only 6 years ago, and ChatGPT was released less than a year ago. It will likely take years, or even decades, before we have a full tech stack for generative AI and LLMs and a host of transformative applications—though we have reason to believe the impact is [accelerated](#) this time around.

I really think that we could be entering a third epoch of computing. The microchip brought the marginal cost of compute to 0. The internet brought the marginal cost of distribution to 0. **These large models actually bring the marginal cost of creation to 0.** When those previous epochs happened, you had no idea what new companies were going to be created. Nobody predicted Amazon. Nobody predicted Yahoo. We should all get ready for a new wave of iconic companies.

—Martin Casado, a16z

We're really at a "Wright-Brothers-first-airplane" moment. We've got something that works and is useful now for some large number of use cases. It looks like it's scaling very, very well, and without any breakthroughs, it's going to get massively better. But **there will be more breakthroughs because now all the AI scientists in the world are working on making**

the stuff better.

—Noam Shazeer, Character.AI

Particularly over the past couple of years and maybe the past 12 months in particular, with the launch of ChatGPT and GPT-4, **you can really see the potential of the platform in the same light as the personal computer or the smartphone. A bundle of technologies are going to make a whole bunch of new things possible that lots of people are going to be building things on top of.**

—Kevin Scott, Microsoft

2. This wave of generative AI has the sort of economics that drive market transformations

For a technological innovation to spur a market transformation, the economics need to be compelling. While previous AI cycles had plenty of technological advances, they lacked transformative economics. In the current AI wave, we have already seen early signs of economics that are 10,000x (or higher) better for some use cases, and the resulting adoption and development of AI appears to be happening much faster than any previous shift.

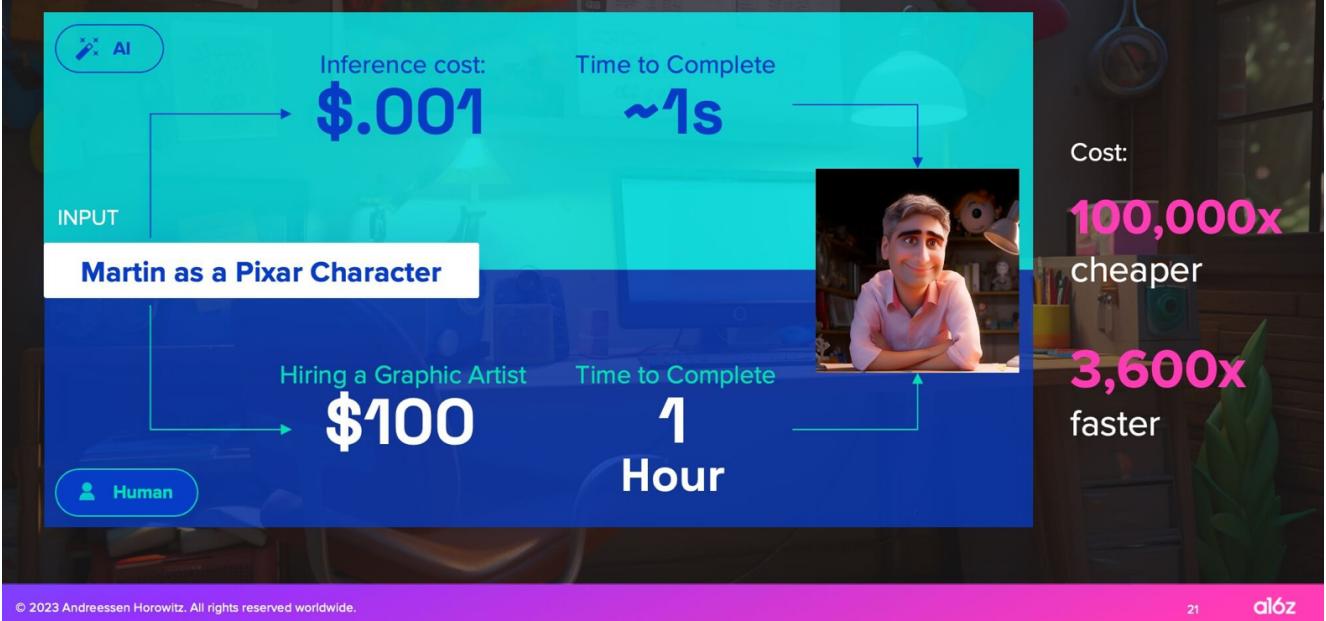
Market transformations aren't created with economics that are 10x better. They get created when they're 10,000x better.

Let's say that I wanted to create an image of myself as a Pixar character. If I'm using one of these image models, the inference costs a 10th of a penny, and let's say it takes 1 second. If you compare that to hiring a graphic artist, let's say that was \$100 in an hour. You've got 4 to 5 orders of magnitude difference in cost and time. These are the types of inflections you look for certainly as an economist when there's going to be actually a massive market dislocation.

And if you want to have an example of how extremely nutty that this can get, I see no reason why you can't generate an entire game—the 3D models, the characters, the voices, the music, the stories, etc. There are companies that are doing all of these things. If you compare the cost of hundreds of millions of dollars and years versus a few dollars, we now have internet and microchip level asymmetries in economics.

—Martin Casado, a16z

Example, Image Creation (Ideogram)



3. For some early use cases, creativity > correctness

Hallucinations are a known aspect of today's LLMs, but for some use cases, the ability to make things up is a feature rather than a bug. In contrast to earlier applied use cases of machine learning where the nth degree of correctness is critical (e.g., self-driving cars), many of the early use cases for LLMs—virtual friends and companions, brainstorming concepts, or building online games—have focused on areas where creativity is more important than correctness..

Entertainment is this \$2T a year industry. And the dirty secret is that entertainment is imaginary friends that don't know you exist. It's a cool first use case for AGI. Like, you want to launch something that's a doctor, it's going to be a lot slower because you wanna be really, really, really careful about not providing false

information. But friends you can do really fast, it's just entertainment, it makes things up, that's a feature.

—Noam Shazeer, *Character.AI*

Of the 65M people on Roblox, most of them are not creating at the level they would want to. For a long time, we imagined a simulation of Project Runway where you have sewing machines and fabrics and it's all 3D simulated, but even that's kind of complex for most of us. I think now when Project Runway shows up on Roblox, it will be a text prompt, image prompt, voice prompt. If I was helping you make that shirt, I'd say: I want a blue denim shirt, I want some buttons, make it a little more trim, fitted. I actually think **we're going to see an acceleration of creation.**

—David Baszucki, *Roblox*

Right now, we're at a place where AI might be doing the first draft, but getting from first draft to final product is kind of hard and usually takes a team. But if you could get AI to suggest interface elements to people and do that in a way that actually makes sense, I think **that could unlock a whole new era of design in terms of creating contextual designs that are responsive to what the user's intent is at that moment.** And I think that'd be a fascinating era for all designers to be working in collaboration with these AI systems.

—Dylan Field, *Figma*

4. For others, like a coding “copilot,” correctness improves as humans use it

While AI has the potential to augment human work across many disciplines, coding “copilots” have been the first AI assistants to see widespread adoption for a few reasons.

First, developers are generally early adopters of new technology—an analysis of ChatGPT prompts from May/June 2023 found that [30% of ChatGPT prompts](#) were coding-related. Secondly, the biggest LLMs have been trained on datasets (e.g., the internet) that contain a lot of code, which makes them particularly good at responding to coding-related queries. Finally, the human in the loop is the user. As a result, while accuracy matters, human developers with an AI copilot can more quickly iterate to correctness than human developers alone.

If you have to be correct and you've got a very long and fat tail of use cases, either you do all of the work technically or you hire people. Often, we hire people. That's a variable cost. Second, because the tails of the solutions tend to be so long—think something like self-driving where there are so many exceptions that could possibly happen—the amount of investment to stay ahead increases and the value decreases. You have this perverse economy of scale. That human in the loop that used to be in a central company is now the user, so it's not a variable cost to the business anymore and the economics for that

work. The human in the loop has moved out, and as a result, you can do things where correctness is important, like, for example, developing code, because it's iterative, so the amount of error that accrues gets less because you're constantly getting feedback and correction from the user.

—Martin Casado, a16z

When a developer can query an AI chatbot to help them write and troubleshoot code, it changes how development happens in two notable ways: 1) it makes it easier for more people to collaborate in development because it happens through a natural language interface and 2) human developers produce more and stay in a state of flow for longer.

Programming is becoming less abstract. We can actually talk to computers in high bandwidth in natural language. We're using the technology and **the technology is helping us understand how to collaborate with it versus 'program' it.**

—Mira Murati, OpenAI

GitHub is the first illustration of this copilot pattern that we are trying to build out, which is: how can you take the knowledge work that someone is doing and use AI to help them be dramatically more productive at doing that particular flavor of cognitive work? In our observation with developers, more than anything else, AI helps keep them in flow state longer than they

otherwise would. Rather than hitting a blocker when you're writing a chunk of code and thinking, "I don't know how to go get the next thing done. I've got to go consult documentation. I've got to go ask another engineer who might be busy with something." **Being able to get yourself unblocked in the moment before you're out of flow state is just extraordinarily valuable.** For folks who are thinking about the utility of these generative AI tools that you're building for things other than software development, this notion of flow state is a useful thing to consider.

—*Kevin Scott, Microsoft*

The best designers are starting to think much more about code, and the best developers are thinking much more about design. Beyond designers and developers, if you think about a product person, for example, they might have been working on a spec before, but now they're going much more into mockups to communicate their ideas more effectively. Basically, this will allow anyone in the organization to go from idea to design and possibly to production much faster. But you'll still need to hone each of those steps. You'll need someone to really think through, "Okay, what ideas are we going to explore? How are we going to explore them?" You'll want to tweak the designs, you'll want to finesse them properly to go from first draft to final product.

—*Dylan Field, Figma*

What comes next?

5. Integrating AI with biology could accelerate new approaches to treating disease and have a profound impact on human health

Biology is incredibly complex—maybe even beyond the ability of the human mind to fully comprehend. The intersection of AI and biology, however, could accelerate our understanding of biology and bring some of the most exciting and transformative technological advances of our time. AI-powered platforms in biology have the potential to unlock previously unknown biological insights, leading to new medical breakthroughs, new diagnostics, and the ability to detect and treat diseases earlier—potentially even stopping a disease before it begins.

At certain times in our history, there have been eras where a particular scientific discipline has made incredible amounts of progress in a relatively short amount of time. In the 1950s, that discipline was computing, where we got these machines that performed calculations that, up until that point, only a human was able to perform. And then in the 1990s, there was this interesting bifurcation. On the one side, there was data science and statistics, which ultimately gave us modern-day machine learning and AI. And then the other side was what I think of as quantitative

biology, which was the first time where we started to measure biology on a scale that was more than tracking 3 genes across an experiment that took 5 years.

Now, this era of 2020 is the time when those last 2 disciplines are actually going to merge and give us **an era of digital biology**, which is the ability to measure biology at unprecedented fidelity and scale; interpret the unbelievable masses of data, different biological scales and different systems using the tools of machine learning and data science; and then bring that understanding back to engineer biology using tools like CRISPR genome editing so **we can make biology do things that it would otherwise not want to do**.

We can now, finally, for the first time, measure biology at scale, both at the cellular level—sometimes at subcellular level—and at the organism level. That gives us, for the very first time, the ability to deploy machine learning in ways where it is truly meaningful.

We built a language model for biology. It's just like GPT, but for cells. We have the language of cells and what cells look like. You measure hundreds of millions of cells in different states, and then just like the large language models for natural language, with a small amount of data, you can start asking, "Okay, how does disease move a disease-causing gene from one place to the other? How does a treatment move you

hopefully back in the opposite direction from the disease state back to the healthy state?" And that's super powerful. Like other language models, it keeps getting better the more data you feed it.

—Daphne Koller, *insitro*

6. Getting models into the hands of users will help us discover new use cases

Whereas previous iterations of AI models were designed to outperform humans at certain tasks, transformer-based LLMs excel at general reasoning. But just because we've created a good general model doesn't mean we've cracked how to apply it to specific use cases. Just as having humans in the loop in the form of RLHF, was critical to improving the performance of today's AI models, getting the new technology into the hands of users and understanding how they use it will be key to figuring out what applications to build on top of those base foundation models.

We have to remember: the model is not the product. Your understanding as an entrepreneur of: who is your user? What is their problem? What can you do to help them? And then determining whether or not AI is actually a useful piece of infrastructure to go solve that user problem—that is something that is unchanged. AI is just like a new and interesting piece of infrastructure that lets you solve a new category of problems or solve an old category of problems in

better ways.

—Kevin Scott, Microsoft

We don't know exactly what the future looks like, and so we are trying to make these tools available and the technology available to a lot of other people so they can experiment and we can see what happens. It's a strategy that we've been using from the beginning.

The week before we launched ChatGPT, we were worried that it wasn't good enough. We all saw what happened: we put it out there and then people told us it is good enough to discover new use cases, and you see all these emergent use cases.

—Mira Murati, OpenAI

7. Your AI friend's memory is about to get much better

While data, compute, and model parameters power the general reasoning of LLMs, context windows power their short-term memory. Context windows are typically measured by the number of tokens that they can process. Today, most context windows are around 32K, but [bigger context windows](#) are coming—and with them, the ability to run larger documents with more context through LLMs.

Right now, the model we're serving uses a context window of a few thousand tokens, which means your lifelong friend remembers what happened for the last half hour. **It will make things way better if you can**

just dump in massive amounts of information. It should be able to know a billion things about you. The HBM bandwidth is there.

—Noam Shazeer, *Character.AI*

One thing that I think is still underappreciated is the longer context and the things that come with that. I think people have this picture in mind of, "There's this chatbot. I ask it a question and it answers the question." But the idea that you can upload a legal contract and say, "what are the 5 most unusual terms in this legal contract?" Or upload a financial statement and say, "summarize the position of this company. What is surprising relative to what this analyst said 2 weeks ago?" **All this knowledge manipulation and processing large bodies of data that take hours for people to read. I think much more is possible with that than what people are doing.** We're just at the beginning of it.

—Dario Amodei, *Anthropic*

8. Voice chatbots, robots, and other ways of interacting with AI are a big area of research

Today, most people interact with AI in the form of chatbots, but that's because chatbots are generally easy to build—not because they're the best interface for every use case.

Many builders are focused on developing new ways for

users to interact with AI models through multimodal AI. Users will be able to interact with multimodal models the way they interact with the rest of the world: through images, text, voice, and other media. A step beyond this: embodied AI focuses on AI that can interact with the physical world, like self-driving cars.

I think that the foundation models today have this great representation of the world in text. **We're adding other modalities, like images and video**, so these models can get a more comprehensive sense of the world, similar to how we understand and observe the world.

—*Mira Murati, OpenAI*

Maybe you want to hear a voice and see a face, or just be able to interact with multiple people. It's like you got elected president, you get the earpiece, and you get the whole cabinet of friends or advisers. Or it's like you walk into Cheers and everyone knows your name, and they're glad you came.

—*Noam Shazeer, Character.AI*

The next frontier of the impacts that AI can have is when AI starts to touch the physical world. And we've all seen just how much harder that is. We've all seen how hard it is astonishingly to build a self-driving car compared to building a chatbot, right? We've made so much progress on building chatbots, and self-driving cars are still blocking fire trucks in San

Francisco. Having an appreciation for that complexity, but also an appreciation for the magnitude of the impact, is important.

—Daphne Koller, insitro

Open Questions

9. Will we have a handful of general models, a bunch of specialized models, or both?

Which use cases are best served by the bigger, “higher IQ” foundation models or smaller, specialized models and datasets? Like the cloud and edge architectural debates of a decade ago, the answer depends on how much you’re willing to pay, how accurate you need the output to be, and how much latency you can tolerate. And the answers to those questions may change over time as researchers develop more compute-efficient methods for fine-tuning big foundation models for specific use cases.

In the long run, we may be over rotating on the question of which models are used for which use case because we are still in the early stages of building the infrastructure and architecture that will support the wave of AI applications to come.

Right now, it’s a bit like 2000 and the internet is about to take over everything, and the most important thing is whoever can build the best router. Cisco in 2000 was worth half a trillion dollars at its peak. It was worth

more than Microsoft at the time. So, who has the largest LLM? Obviously, whoever can build the largest one and train it the most will own all of AI and all future humanity. But just like the internet, someone will show up later and think about something like Uber and cab driving. Someone else showed up and thought, "hey, I wanna check out my friends on Facebook." Those end up being huge businesses, and **it's not just going to be one model that OpenAI or Databricks or Anthropic or someone builds, and that model will dominate all these use cases.** A lot of things will need to go into building the doctor that you trust.

—Ali Ghodsi, *Databricks*

The biggest factor is simply that more money is being poured into it. The most expensive models made today cost about \$100M, say, plus or minus a factor of 2. **Next year we're probably going to see, from multiple players, models on the order of \$1B**, and in 2025, we're going to see models on the order of several billion, perhaps even \$10B. That factor of 100, plus the compute inherently getting faster with the H100s—that's a particularly big jump because of the move to lower precision. You put all those things together, and if the scaling laws continue, there's going to be a huge increase in capabilities.

—Dario Amodei, *Anthropic*

It depends on what you're trying to do. Obviously, the trajectory is these AI systems will be doing more

and more of the work that we're doing. In terms of how this works out with the OpenAI platform, you can see even today that we make a lot of models available through our API, from the very small models to our frontier models. People don't always need to use the most powerful or the most capable models. Sometimes they just need the model that actually fits for their specific use case and it's far more economical. We want people to build on top of our models and we want to give them tools to make that easy. We want to give them more and more access and control, so you can bring your data and customize these models. And you can really focus on the layer beyond the model and defining the products.

—*Mira Murati, OpenAI*

In any company like Roblox, there's probably 20 or 30 ultimately end-user vertical applications that are very bespoke—natural language filtering is very different from generative 3D—and at the end user point, we want all of those running. As we go down [the stack], there's probably a natural 2 or 3 clustering of general bigger, fatter type models in a company like ours. We're very fine-tuned for the disciplines we want with the ability to train and run massive inferences for these.

—*David Baszucki, Roblox*

10. When will AI gain traction in the

enterprise, and what will happen to those datasets?

The impact of generative AI on the enterprise is still in its infancy – in part because enterprises tend to move slower generally, and in part because they have realized the value of their proprietary datasets and don't necessarily want to hand their data over to another company, no matter how powerful their model might be. Most enterprise use cases require a high degree of accuracy, and enterprises have 3 options for choosing an LLM: build their own, use an LLM service provider (e.g., Mosaic) to build it for them, or fine tune a foundation model – and building your own LLM is not easy

One of the things that's happened in the brains of the CEOs and the boards is that they realize: maybe I can beat my competition. Maybe this is the kryptonite to kill my enemy. I have the data with generative AI, so then they're thinking, 'I have to build it myself.' I have to own the IP. **You want to build your own LLM from scratch? It's not for the faint of heart, still requires a lot of GPUs, costs a lot of money, and depends on your datasets and your use cases.**

—Ali Ghosdi, *Databricks*

We have lots and lots of customers who want to have specialized models that are cheaper, smaller, and have really high accuracy and performance.
They're saying, "Hey, this is what I want to do. I want to

classify this particular defect in the manufacturing process from these pictures really well." And there, the accuracy matters. Every ounce of accuracy that you can give me matters. And there, you're better off if you have a good dataset to train and can train a smaller model. The latency will be faster, it will be cheaper, and yes, you can absolutely have accuracy that beats the really large model. But that model that you built can't also entertain you on the weekend and help your kids do their homework.

—Ali Ghodsi, *Databricks*

11. Will scaling laws take us all the way to AGI?

LLMs currently follow a scaling law: model performance improves as you add more data and compute, even if the architecture and algorithms remain constant. But how long will this scaling law hold? Will it continue indefinitely, or will it hit a natural limit before we develop AGI?

There isn't any evidence that we will not get much better and much more capable models as we continue to scale them across the access of data and compute. Whether that takes you all the way to AGI or not—that's a different question. **There are probably some other breakthroughs and advancements needed along the way, but I think there's still a long way to go in the scaling laws** and to really gather a lot of benefits from these larger models. —Mira Murati, *OpenAI*

Even if there were no algorithmic improvements from here and we just scaled up what we had so far, **the scaling laws are going to continue.** —Dario Amodei, *Anthropic*

Our goal is to be an AGI company and a product-first company, and the way to do that is by picking the right product that forces us to work on things that generalize, make the model smarter, make it what people want, and serve it cheaply at a massive scale. **The scaling laws are going to take us a pretty long way. At the core, the computation isn't that expensive.** Operations cost something like 10-18 dollars these days. If you can do this stuff efficiently, the cost of that should be way, way lower than the value of your time. There's the capacity there to scale these things up by orders of magnitude. —Noam Shazeer, *Character.AI*

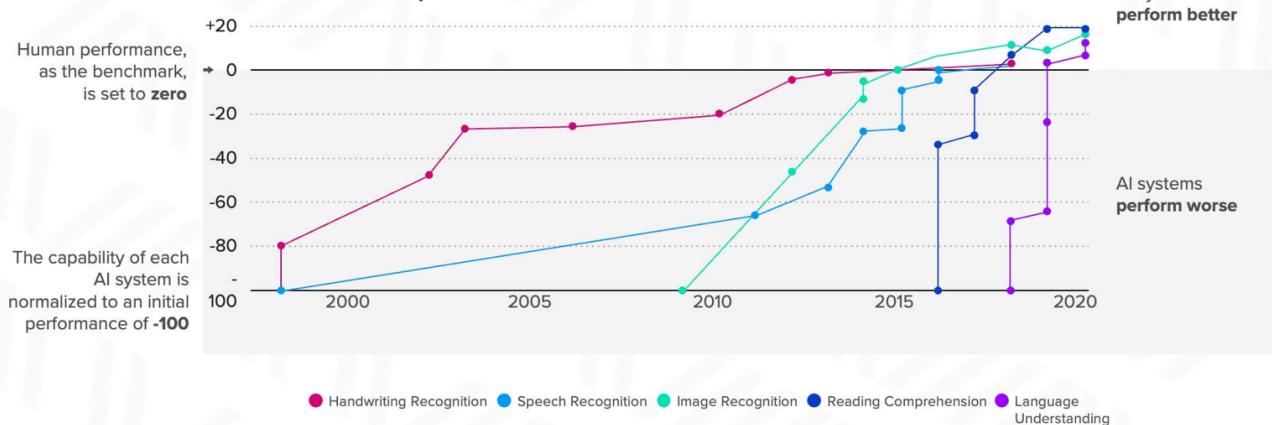
12. What are the emergent capabilities?

While some people are quick to write off generative AI's capabilities, AI is already much better than humans at performing certain tasks and will continue to improve. The best builders have been able to identify AI's most promising emerging capabilities and build models and companies to scale those capabilities into reliable features. They recognize that scale has tended to increase the reliability of emergent abilities.

AI is Better than Humans at Lots of Things

Language and image recognition capabilities of AI Systems have improved rapidly

Test scores of the AI relative to human performance:



Data Source: Kiel et al. (2021) - Dynabench: Rethinking Benchmarking in NLP
OurWorldinData.org - Research and data to make progress against the world's largest problems.

© 2023 Andreessen Horowitz. All rights reserved worldwide.

6

ai6z

It's important to pay attention to these emergent capabilities even if they're highly unreliable.

Especially for people that are building companies today, you really want to think about, "Okay, what's somewhat possible today? What do you see glimpses of today?" Because very quickly, these models could become reliable.

—Mira Murati, OpenAI

When we put out GPT-2, some of the stuff that was considered most impressive at the time was, "You put these 5 examples of English to French translation straight into the language model, and then you put a sixth sentence in English and it actually translates into French. Like, oh, my God, it actually understands the pattern." That was crazy to us, even though the translation was terrible. But our view was that, "Look, this is the beginning of something amazing because there's no limit and you can continue to scale it up."

There's no reason why the patterns we've seen before won't continue to hold. **The objective of predicting the next word is so rich, and there's so much you can push against that it just absolutely has to work.** And then some people looked at it and they were like, "You made a bot that translates really badly." —Dario Amodei, Anthropic

13. Will the cost of serving these models come down?

The cost of compute is one of the major constraints to scaling these models, and the current chip shortages drive that cost up by constraining supply. However, if Nvidia produces more H100s next year, [as some have reported](#), that should ease the GPU shortage and, with it, potentially the cost of compute.

The model we're serving now cost us about \$2M worth of compute cycles to train last year, and we could probably repeat it for half a million now. So we're going to launch something tens of IQ points smarter, hopefully, by the end of the year. **I see this stuff massively scaling up. It's just not that expensive.** I think I saw an article yesterday that NVIDIA is going to build another 1.5M H100s next year. That's 2M H100s, so that's $2 \times 10^6 \times 10^{15}$ operations per second. So, $2 \times 10^{21} / [8 \times 10^9]$ people on Earth.

So that's roughly a quarter of a trillion operations per

second, per person. Which means it could be processing on the order of 1 word per second on a 100B parameter model for everyone on Earth. But really it's not going to be everyone on Earth because some people are blocked in China, and some people are sleeping... It's not that expensive. This thing is massively scalable if you do it right, and we're working on that.

—Noam Shazeer, Character.AI

My basic view is that inference will not get that much more expensive. The basic logic of the scaling laws is that if you increase compute by a factor of n, you need to increase data by a factor of the square root of n and the size of the model by a factor of square root of n. That square root basically means that the model itself does not get that much bigger, and the hardware is getting faster while you're doing it. **I think these things are going to continue to be serve-able for the next 3 or 4 years. If there's no architectural innovation, they'll get a little bit more expensive. If there's architectural innovation, which I expect there to be, they'll get somewhat cheaper.**

—Dario Amodei, Anthropic

But even if the cost of compute remains the same, efficiency gains at the model level seem inevitable, especially with so much talent flooding the space—and AI itself is likely to be the most powerful tool we have to improve how AI works.

As AI gets more powerful, it gets better at most cognitive tasks. One of the relevant cognitive tasks is judging the safety of AI systems, eventually doing safety research. There's this self-referential component to it. We see it with areas like interpretability, looking inside the neural nets.

Powerful AI systems can help us interpret the neurons of weaker AI systems. And those interpretability insights often tell us a bit about how models work. And when they tell us how models work, they often suggest ways that those models could be better or more efficient. —*Dario Amodei, Anthropic*

One of the most promising areas of research is fine-tuning a big model for a specific use case without needing to run the entire model.

If you made a thousand versions of an LLM, that's good at a thousand different things, and you have to load each of those into the GPUs and serve them, it becomes very expensive. **The big holy grail right now that everybody's looking for is: are there techniques, where you can just do small modifications where you can get really good results?** There are lots of techniques, like prefix tuning, LoRA, CUBE LoRA, etc. But jury's out, none of them really are a slam dunk, awesome, we found it. But someone will.

Once you have that, then it seems the ideal would be a really big foundation model that's pretty smart and that you can stack on these additional brains that are really good at, say, this specific classification task for manufacturing errors.

—Ali Ghodsi, Databricks

14. How do we measure progress towards AGI?

As we scale these models, how will we know when AI becomes AGI? As often as we hear the term AGI, it can be a tricky thing to define, perhaps in part because it's difficult to measure.

Quantitative benchmarks like [GLUE](#) and [SUPERGLUE](#) have long been used as standardized measures of AI model performance. But as with the standardized tests we give to humans, AI benchmarks raise the question: to what extent are you measuring an LLM's reasoning capabilities and to what extent are you measuring its ability to take a test?

I kind of think all the benchmarks are bullshit.

Imagine if all our universities said, "we're going to give you the exam the night before and you can look at the answers. And then the next day, we're going to bring you in and you answer them, and then we'll score how you did." Suddenly, everybody would be acing their exams.

For instance, MMLU is what a lot of people benchmark these models on. MMLU is just a multi-choice question that's on the web. Ask a question, and is the answer A, B, C, D, E? And then it says what the right answer is. It's on the web. You can train on it and create an LLM that crushes it on that.

—Ali Ghodsi, Databricks

The original qualitative test for AGI was the Turing Test, but convincing humans that an AI is a human is not the hard problem. Getting an AI to do what humans do in the real world is the hard problem. So, what tests can we use to understand what these systems are capable of?

We're now seeing from these systems that it's easy to convince a human that you're human, but it's hard to actually make good things. I could have GPT-4 create a business plan and come pitch to you, but that doesn't mean you're going to invest. When you actually have 2 competing businesses side-by-side—one of them is run by AI and the other is run by a human—and you invest in the AI business, then I'm worried.

—Dylan Field, Figma

I have a Turing test question for AI: if we took AI in 1633 and trained on all the available information at that time, would it predict that the Earth or the sun is the center of the solar system—even though 99.9% of the information is saying the Earth is the center of the solar

system? I think 5 years is right at the fringe, but if we were to run that AI Turing test 10 years from now, it would say the sun.

—*David Baszucki, Roblox*

15. Will we even need humans in the loop?

New technologies have typically replaced some human work and jobs, but they've also opened entirely new fields, improved productivity, and made more types of work accessible to more people. While it's easy to imagine AI automating existing jobs, it's far more difficult to imagine the next order problems and possibilities unlocked by AI.

Very simply, Jevons Paradox states: if the demand is elastic and the price drops, the demand will more than make up for it. Normally, far more than make up for it. This is absolutely the case of the internet. You get more value and more productivity. I personally believe when it comes to any creative asset or work automation, the demand is elastic. The more that we make, the more people consume. We're very much looking forward to a massive expansion in productivity, a lot of new jobs, a lot of new things, just like we saw with the microchip and the internet.

—*Martin Casado, a16z*

I grew up in rural Central Virginia, in a place where the economy was powered by tobacco farming, furniture manufacturing, and textiles. By the time I was

graduating from high school, all 3 of those industries had just collapsed. When folks in these communities have access to very powerful tools, they tend to do remarkable things that create economic opportunities for themselves, their families, and their communities. They solve problems that you or I are not going to solve just because we don't see the whole problem landscape of the world. We don't have their point of view. These tools of AI are becoming more accessible now than they have been by a dramatic wide margin. You can do interesting things with these tools right now and can be an entrepreneur in small-town Virginia without having a PhD in computer science or expertise in classical AI. You just need to be curious and entrepreneurial.

—Kevin Scott, Microsoft

If you look at every technological shift or platform shift so far, it's resulted in more things to design. You got the printing press, and then you have to figure out what you put on a page. More recently, mobile. You would think, "Okay. Less pixels, less designers." But no, that's when we saw the biggest explosion of designers.

—Dylan Field, Figma

Calling all builders

16. There's never been a more exciting time to build an AI startup (especially if you're a physicist or mathematician)

It's a uniquely exciting time to build in AI: foundation models are rapidly scaling, the economics are finally tilting in startups' favor, and there are plenty of problems to solve. These problems will require a lot of patience and tenacity to crack, and to date, physicists and mathematicians have been particularly well suited to working on them. But as a young field that's also advancing quickly, AI is wide open—and there's never been a better time to build in AI.

There's two kinds of fields at any given point in time. There's fields where an enormous amount of experience and accumulated knowledge has been built up and you need many years to become an expert. The canonical example of that would be biology—it's very hard to contribute groundbreaking or Nobel Prize-level work in biology if you've only been a biologist for 6 months... Then there are fields that are very young or that are moving very fast. **AI was, and still is to some extent, very young and is definitely moving very fast. Really talented generalists can often outperform those who have been in the field for a long time because things are being shaken up so much. If anything, having a lot of prior knowledge can be a disadvantage.**

—Dario Amodei, Anthropic

One thing to draw on from the theoretical space of math is that you need to sit with a problem for a really long time. Sometimes you sleep and you wake

up and you have a new idea, and over the course of a few days or weeks, you get to the final solution. It's not a quick reward, and sometimes it's not this iterative thing. It's almost a different way of thinking where you're building the intuition and discipline to sit with a problem and have faith that you're going to solve it. Over time, you build an intuition on what problem is the right problem to actually work on.

—*Mira Murati, OpenAI*

It's not only machine learning that moves forward over time, it's also the biological tools that we're relying on. It used to be that there wasn't any CRISPR. There was just siRNA. And then there's CRISPR base editing and now there's CRISPR prime that replaces entire regions of the genome. So the tools that we're building on also get better and better over time, which unlocks more and more diseases that we could tackle in a meaningful way. **There's so many opportunities at this intersection of AI/ML and biology and medicine. This convergence is a moment in time for us to make a really big difference in the world that we live in using tools that exist today that did not exist even five years ago.**

—*Daphne Koller, insitro*

If you think about some of these big platform shifts that we've had in the past, the most valuable things that get done on the platforms are not the things that got deployed in the first year or two of the platform

change. If you think about where you spend most of your time on your smartphone, it's not the SMS app, it's not the web browser, it's not the mail client. It's the new things that got created on top of the platform in the years following the availability of the platform.

What are the difficult things that have now become possible that were impossible before? That's the thing that people ought to be thinking about. Don't chase the trivial things.

—Kevin Scott, Microsoft

Visit a16z.com/AIRevolution and [YouTube](#), for the full series, including:

[The Economic Case for Generative AI](#) with a16z's Martin Casado

[Where We Go from Here](#) with OpenAI's Mira Murati

[Leveling Up](#) with Roblox's David Baszucki

[Democratizing Design](#) with Figma's Dylan Field

[Digital Biology](#) with insitro's Daphne Koller

[Improving AI](#) with Anthropic's Dario Amodei

[AI Copilots and the Future of Knowledge Work](#) with Microsoft's Kevin Scott

[AI Food Fights in the Enterprise](#) with Databricks' Ali Ghodsi

[Universally Accessible Intelligence](#) with Character.ai's Noam Shazeer

Conversations and quotes have been condensed and

edited for readability and clarity.